

Key words: coronavirus MHV|sequence|translation initiation|hydrophobicity

## Coronavirus MHV-JHM mRNA 5 Has a Sequence Arrangement which Potentially Allows Translation of a Second, Downstream Open Reading Frame

By M. A. SKINNER,\*† D. EBNER AND S. G. SIDDELL

*Institute of Virology, Versbacher Strasse 7, 8700 Würzburg, Federal Republic of Germany*

(Accepted 20 November 1984)

### SUMMARY

The sequence of a 5'-proximal region of mRNA 5 of coronavirus MHV-JHM was determined by chain-terminator sequencing of cDNA subcloned in M13. The sequence contained two long open reading frames of 321 bases and 264 bases, overlapping by five bases but in different frames. Both open reading frames are initiated by AUG codons in sequence contexts that are relatively infrequently used as initiator codons. The smaller, downstream open reading frame encoded a neutral protein (mol. wt. 10200) with a hydrophobic amino terminus. The larger, 5'-proximal open reading frame encoded a basic protein (mol. wt. 12400) which lacks internal methionine residues. With the exception of the AUG codon initiating the downstream open reading frame, no internal AUG codons were found within the sequence covered by the upstream open reading frame. These results suggest that the MHV-JHM mRNA 5 is translated to produce two proteins by a mechanism involving internal initiation of protein synthesis. Preliminary evidence is presented showing that the downstream open reading frame is functional *in vivo*.

### INTRODUCTION

Coronaviruses are pleomorphic, enveloped viruses which replicate in the cytoplasm of vertebrate cells. Their genome is a single-stranded, infectious RNA of mol. wt. about  $6 \times 10^6$ . Their molecular biology has recently been reviewed by Siddell *et al.* (1983). The most widely studied member of the coronavirus group is murine hepatitis virus (MHV). MHV virions contain three structural proteins: peplomer (or E2), membrane (or E1) and nucleocapsid protein. Infection by MHV results in the production of seven mRNA species in infected cells, representing the genomic RNA and six subgenomic mRNAs (mol. wt. from  $0.6 \times 10^6$  to  $3.7 \times 10^6$ ). The mRNAs form a nested set with a common 3' terminus. Lai *et al.* (1983) showed that each mRNA possesses a common 5' leader, derived from the 5' end of the genomic RNA. Sequencing of mRNAs 6 and 7 (Armstrong *et al.*, 1984; Skinner & Siddell, 1983) and of an intergenic region of genomic RNA (Spaan *et al.*, 1983) showed that this leader is about 70 bases long.

The translation *in vitro* of size-fractionated MHV mRNAs in cell-free systems or oocytes (Rottier *et al.*, 1981; Leibowitz *et al.*, 1982; Siddell, 1983) has shown that the major primary translation products of mRNAs 3, 6 and 7 are the polypeptide components of the virion peplomer, membrane and nucleocapsid proteins respectively. The size of the primary translation products (150K, 26K, 50K) and the size of the 'unique' sequences in the respective mRNAs (4.5 kb, 0.7 kb, 1.8 kb) suggests that the unique sequences encode and are translated into a single polypeptide. The translation *in vitro* of the genome-sized mRNA 1 to produce a series of related, approximately 200K polypeptides, which are thought to represent viral polymerase components (Leibowitz *et al.*, 1982), together with sequence analysis of mRNAs 6 and 7 (Armstrong *et al.*, 1984; Skinner & Siddell, 1983) are also consistent with this idea. Only

† Present address: Department of Microbiology, University of Leicester, Leicester LE1 7RH, U.K.

mRNA 2 which translates *in vitro* to produce a 30K to 35K polypeptide but has a 'unique' coding capacity of 80K is not fully consistent with such a model.

The MHV subgenomic mRNAs are produced in non-equimolar amounts in infected cells. Messenger RNAs 4 and 5 are minor species and have similar sizes. The translation *in vitro* of RNA fractions containing both mRNAs has produced a 14K to 14.5K viral polypeptide (Leibowitz *et al.*, 1982; Siddell, 1983) but assignment to a specific mRNA has not been possible. In the accompanying paper (Skinner & Siddell, 1985), we present evidence that the RNA sequences comprising the unique region of mRNA 4 encode and are translated to produce this polypeptide. In this paper we present the sequence for the unique region of mRNA 5. Unexpectedly, our analysis suggests that this region, in contrast to the other MHV mRNAs (with the possible exception of mRNA 2), encodes two proteins. The organization of the coding sequences and the implications of these results are discussed.

## METHODS

**Materials.** Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences (St Petersburg, Fla., U.S.A.). Synthetic oligonucleotides for cDNA synthesis, M13 sequencing (17-mer), M13 hybridization probes, as well as oligo(dG)<sub>12-18</sub> were supplied by Pharmacia P-L Biochemicals. *Escherichia coli* DNA polymerase I, S1 nuclease, terminal deoxynucleotidyl transferase and T4 polynucleotide kinase were obtained from Bethesda Research Laboratories. Lyophilized calf intestinal alkaline phosphatase was from Boehringer Mannheim. T4 DNA ligase was from New England Nuclear. Restriction enzymes were from Pharmacia P-L Biochemicals, Bethesda Research Laboratories, Boehringer Mannheim and Renner (Dannstadt, F.R.G.). Radiochemicals were supplied by Amersham Buchler.

**Synthesis and cloning of double-stranded cDNA.** Isolation of virus and of polyadenylated RNA from MHV-infected Sac(-) cells was performed as described previously (Siddell *et al.*, 1980). Genomic RNA was isolated by phenol/chloroform extraction of purified virus. Single-stranded cDNA was prepared according to protocols described by Land *et al.* (1983) except that a synthetic primer (3'-ATTAGATTTGA-5', Pharmacia P-L Biochemicals) was used, at a concentration of 300 µg/ml, instead of oligo(dT). The primer is complementary to genomic and mRNA 7 sequences just upstream of the initiation site for translation of the nucleocapsid protein (Skinner & Siddell, 1983). Second-strand cDNA synthesis, cloning of double-stranded cDNA and characterization of cloned cDNA were as previously described (Skinner & Siddell, 1983), except that mapping of restriction enzyme sites was not performed (sizes of restriction fragments were, however, determined).

**Nucleotide sequencing.** Fragments of cDNA inserts were generated by a variety of restriction enzymes and were either cloned as a mixture or as single fragments (purified by electroelution from polyacrylamide gels) into the M13 vectors mp8 and mp9 (Messing & Vieira, 1982). Fragments were then sequenced using the chain-terminator method of Sanger *et al.* (1977). Sequence data were analysed and assembled by the programs of Staden (1982).

**Direct sequencing of RNA.** A 56 bp fragment of DNA was isolated following cleavage of the MHV-A59-specific cDNA clone (in pAGS1125) with *AluI* (positions 297 to 352 in Fig. 3). The fragment (400 ng) was annealed to 60 µg of MHV-A59-infected cell RNA in 80% formamide, 40 mM-PIPES pH 6.4, 1 mM-EDTA, 0.4 M-NaCl at 37 °C for 3 h (the melting temperature of the fragment in this buffer having been determined to be 32 to 34 °C). The annealed RNA and primer were precipitated in 0.3 M-sodium acetate, 70% ethanol and chain-terminator sequencing was performed with 10 µg of the annealed RNA and primer in 50 mM-Tris-HCl pH 8.3, 50 mM-KCl, 8 mM-MgCl<sub>2</sub>, 1 mM-dithiothreitol and 1 unit RNase inhibitor (Amersham Buchler) per µl. Each 10 µl reaction contained 3 units reverse transcriptase, 20 µCi [ $\alpha$ -<sup>32</sup>P]dATP (3000 Ci/mmol), and 7 pmol dATP. Dideoxy ATP was used at 0.2 µM, ddCTP, ddGTP and ddTTP were used at 2.5 µM, dCTP, dGTP and TTP were at 25 µM. After 30 min at 42 °C, a chase was performed with 50 µM-dNTP for 30 min at 42 °C. Electrophoresis was as described by Sanger *et al.* (1977).

**Primer extension on infected cell mRNA.** The same primer as used for cloning was dephosphorylated and 5' end-labelled with <sup>32</sup>P using protocols described by Maniatis *et al.* (1982). Two pmol of the primer and 6 µg poly(A)<sup>+</sup> RNA from MHV-A59-infected cells were heated at 95 °C for 3 min, frozen on dry ice and then thawed in 50 mM-Tris-HCl pH 8.3, 140 mM-KCl, 8 mM-MgCl<sub>2</sub>, 4 mM-sodium pyrophosphate, 0.4 mM-dithiothreitol, 1 mM-dNTPs and 1 unit RNase inhibitor per µl. Reverse transcriptase (50 units) was added and the reaction was incubated at 42 °C for 1 h when a further 50 units of reverse transcriptase was added and incubation was continued for another hour. Following phenol extraction and ethanol precipitation, a quarter of the sample was electrophoresed by alkaline agarose electrophoresis (McDonnell *et al.*, 1977) in a vertical gel. The gel was neutralized in 7% TCA (30 min) and after drying was exposed to Fuji RX film (without screens).

**Northern hybridizations.** Infected cell RNA was electrophoresed in 1% formaldehyde-agarose gels and was transferred onto nitrocellulose (Schleicher & Schüll) according to Maniatis *et al.* (1982).

M13 hybridization probes were made from sequenced M13 clones of the appropriate polarity using the method of Hu & Messing (1982). The 56 bp *AluI* fragment described earlier was labelled at the 5' end using protocols described by Maniatis *et al.* (1982). Cloned mRNA 7 cDNA (Skinner & Siddell, 1983) was labelled by nick translation using the method of Rigby *et al.* (1977).

Prehybridization was carried out in 50% formamide, 5 × SSPE, 5 × Denhardt's solution, 100 µg denatured salmon sperm DNA per ml, 0.1% SDS at 42 °C and hybridizations were performed in the same buffer except that it contained 1 × Denhardt's solution. Filters were washed twice in 2 × SSPE, 0.1% SDS and twice in 0.1 × SSPE, 0.1% SDS, at room temperature.

*Labelling and electrophoresis of intracellular proteins.* Procedures for the infection, labelling and preparation of total or cytoplasmic cell lysates of MHV-infected cells or mock-infected Sac(−) have been described previously (Siddell *et al.*, 1980, 1981). Samples were electrophoresed on 15% discontinuous SDS-polyacrylamide gels as described by Laemmli (1970).

*Translation in vitro of size-fractionated RNA.* Cytoplasmic, polyadenylated RNA from MHV-JHM-infected cells was fractionated on sucrose-formamide gradients and was translated in an L-cell lysate as described previously (Siddell, 1983).

## RESULTS

### *Identification of open reading frames*

MHV-JHM-specific cDNA was synthesized using intracellular polyadenylated RNA as a template. The largest cDNA clone isolated, in pJMS1010, hybridized against all the viral mRNAs except mRNA 7. Sequence determination and comparison with the MHV-A59 sequence for mRNA 6 (Armstrong *et al.*, 1984; M. A. Skinner, unpublished) revealed that one end of the clone was positioned 297 bases upstream from the priming sequence, possibly due to incomplete second-strand synthesis. This position, 1983 bases from the 3' end of the genome, excluding the poly(A) tail, is designated as −1983. The beginning of the membrane protein E1-coding sequence (encoded by mRNA 6) was identified at position −2370 and three large open reading frames (ORF) were found at positions −2633 to −2370 (ORF C), −2949 to −2629 (ORF B) and −3350 to −2934 (ORF A). Upstream of ORF A is an ORF of 1160 bases (M. A. Skinner, unpublished) extending up to (and presumably beyond) the end of the currently sequenced DNA. Fig. 4(a) shows the general arrangement of these ORFs.

As the sizes reported for MHV mRNAs 4 and 5 vary considerably (from  $1.2 \times 10^6$  to  $1.5 \times 10^6$  for mRNA 4 and from  $1.08 \times 10^6$  to  $1.2 \times 10^6$  for mRNA 5; see review by Siddell *et al.*, 1982), we decided to map these ORFs to the mRNAs by hybridization analysis. A 56 base pair *AluI* restriction fragment from ORF B, an M13 hybridization probe from ORF A and an M13 hybridization probe from a position upstream of ORF A (within the 1160 base ORF) were hybridized against a nitrocellulose filter to which viral mRNAs, fractionated by formaldehyde-agarose electrophoresis, had been transferred (for exact positions of these probes see Fig. 1). This analysis (Fig. 1) showed that the region 560 bases and more upstream of ORF A (and therefore within the 1160 base ORF) was located in the 'unique' sequence of mRNA 3. ORF A was located in the 'unique' region of mRNA 4 and ORF B (and, therefore, also ORF C) was within the 'unique' region of mRNA 5.

We then used primer extension analysis to map the 5' ends of the mRNAs. The primer used for cloning was extended on infected cell RNA (Fig. 2) and extension products were assigned to the subgenomic mRNAs on the basis of the relative intensity of the stops (compared to the relative abundance of mRNAs in MHV-A59-infected cells, Fig. 1), the approximate sizes of the mRNAs and the hybridization data described above. The two strongest stops, at about 80 and 800 bases, corresponded well to the known sizes of mRNA 7 and mRNA 6 (showing them to end at −1755 and −2475, respectively). A clear, but fainter stop (at about 1400 bases, −3075) was assigned to mRNA 5. A number of minor extension products (about 1200 bases) were observed, but these corresponded well with a run of stops observed during direct chain-terminator sequencing of MHV-A59-specific intracellular RNA (see below and Fig. 3) and most likely do not represent mRNA termini. Finally, a clear and yet fainter stop (at about 1800 bases, −3475) was assigned to mRNA 4. Allowing for a leader sequence of about 70 bases at the 5' end of each mRNA, this result suggests that the body of mRNA 4 begins at about position −3400 and the body of mRNA 5 begins at about position −3000.

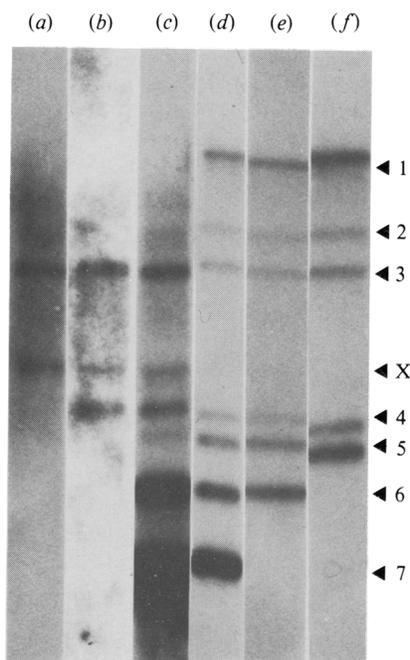


Fig. 1

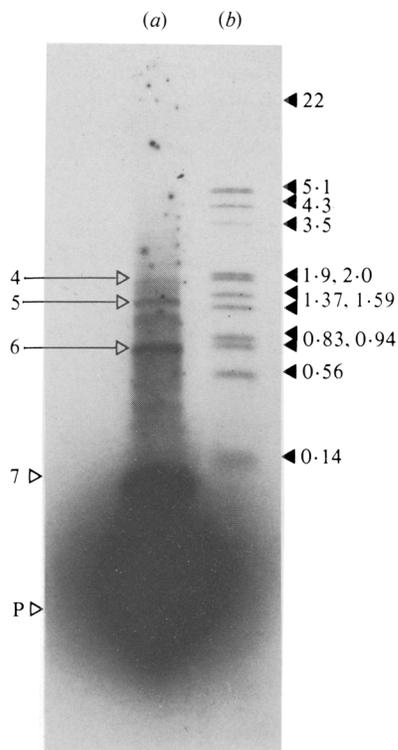


Fig. 2

Fig. 1. Hybridization of M13 probes and an *AluI* restriction fragment to infected cell RNA separated by formaldehyde-agarose electrophoresis and transferred to nitrocellulose. (a to c) MHV-JHM-infected cell RNA; (d to f) MHV-A59-infected cell RNA. Probes used were: (a) M13 probe from 554 to 947 bases upstream of ORF A; (b) M13 probe from position -3137 to -3060, within ORF A; (c, d) nick-translated mRNA 7 cDNA clone (Skinner & Siddell, 1983); (e) nick-translated MHV-A59 cDNA clone (pAGS1125) extending from the primer (-1676) to -2840 and therefore including the ORF for membrane protein E1; (f) kinase-labelled *AluI* fragment (-2731 to -2676) from within ORF B. Numbers refer to mRNAs. X is a minor RNA species, the relative abundance of which varies from preparation to preparation. Its nature is currently under investigation.

Fig. 2. Primer extension on infected cell RNA to map the 5' ends of subgenomic mRNAs. The synthetic primer used for cloning was extended on poly(A)<sup>+</sup> RNA isolated from cells infected with MHV-A59 as described in Methods. The products were analysed on a 1% vertical alkaline agarose gel. (a) Primer extension; extension products assigned to the 5' ends of subgenomic mRNAs are indicated. P represents the position of the 11-mer primer. (b) Labelled DNA markers ( $\lambda$  *EcoRI/HindIII* digest); lengths in kilobases.

#### 5' non-coding sequences of mRNA 5

Upstream of the apparent coding region of mRNA 5 we have been able to identify the sequence GUUCUAAAC. This sequence is very similar to the sequence AAUCUAAAC which is found upstream of the mRNA 4-coding sequence (Skinner & Siddell, 1985). The latter sequence is identical to a sequence in the intergenic region upstream of the coding sequence of mRNA 7 (Spaan *et al.*, 1983) and differs by only one base from a sequence upstream of the E1-coding sequence of mRNA 6 (AAUCCAAAC; M. A. Skinner, unpublished). It was postulated that such homologous sequences might be involved in regulating the initiation of synthesis of the bodies of MHV mRNAs (Armstrong *et al.*, 1983; Spaan *et al.*, 1983).

```

1 TAGTCTAAACCTCATCTTAAATTTCTGGCCGTCACATACACACTTAGGCACCTTGCCGAAGTA 60

MetThrProProAlaThrTrp..Ile----
TATGAGACCAACNGCCACATGGXXATTTGGCATGTNAGTGATGCNTGGTTNCGCCGCNCG
61 TATGACACCACCAGCTACATGGAGATTGGCTTGTGAGTACGCCTGGCTACGCCGAACG 120
MetGluIleTrpLeuValSerAspAlaTrpLeuArgArgThr

CGGGACTTTGGTGTCNNTCGCCTNGAAGATNNNNNNNNNNNNNATTAJAGCCAACCC
121 CGAGACTTTGGTGCTCACTCGACTTGAAGATTTTGGCTTCCAATTAATTATTGCCAACCC 180
ArgAspPheGlyValThrArgLeuGluAspPheCysPheGlnPheAsnTyrCysGlnPro

CGAGTNN/
GTTATTGTAGAGTTCCTTTAAAGGCTTGGGTAGCAACCGGGTAAATTTGCA
181 CGAGTTGGTTATTGTAGAGTTCCTTTAAAGGCTTGGGTAGCAACCGGGTAAATTTGCA 240
ArgValGlyTyrCysArgValProLeuLysAlaTrpCysSerAsnGlnGlyLysPheAla

GCGCAGTTTACCCTAAAAAGTTGCGAAAAACCGGTACGAAAAATTTACTAGCTTC
241 GCGCAGTTTACTCTTAAAAAGTTGCGAAAAATCAGGCCACAAAAATTCATTTACTAGCTTC 300
AlaGlnPheThrLeuLysSerCysGluLysSerGlyHisGlnLysPheIleThrSerPhe

ACGGCCTACGGCAGAAGTGTCCAACAGGCCGTTAGCAAGTTAGTAGAAGAAGCTGTTGAT
301 ACGGCCTACGGCAGAAGTGTCCAACAGGCCGTTAGTAAGCTAGTAGAAGAAGCTGCTGAT 360
ThrAlaTyrAlaLysThrValLysGlnAlaValSerLysLeuValGluGluAlaAlaAsp

TTTATTGTTTTTAGGGCCACGCAGCTCGAAAGAAATGTTTAAATTTATTCTTTACAGACAC
361 TTTATCATCTGGAGAGCCACGCAGCTCGAAAGAAATGTTTAAATTTATTCTTTACAGACAC 420
PheIleIleTrpArgAlaThrGlnLeuGluArgAsnValEnd
MetPheAsnLeuPheLeuThrAspTh

AGTATGGTATGTGGGGCAGATTATTTTATATTCGCAGTGTGTTGATGGTCACCATAAT
421 AGTATGGTATGTGGGGCAGATTATCTTTATAGTCGCAGTGTGTTGATGGTCACCATAAT 480
rValTrpTyrValGlyGlnIleIlePheIleValAlaValCysLeuMetValThrIleIleI

TGTGGTGCCTTCCTTGCCTCTATCAAACCTTGTATTCAAACCTTGCGGTTTATGTAATACT
481 TGTGGTGCCTTCCTTGCCTCTATTAACGTTGTATTCAAACCTTGCGGTTTATGTAATACT 540
eValValAlaPheLeuAlaSerIleLysArgCysIleGlnLeuCysGlyLeuCysAsnTh

TTTGGTGTCTCCCTTCTATTTATTTGTATGATAGGAGTAAGCAGCTTTATAAGTATTA
541 TTTGGTGTCTCCCTTCTATTTATCTGTATAATAGGAGTAAGCAGCTTTATAAGTATTA 600
rLeuLeuLeuSerProSerIleTyrLeuTyrAsnArgSerLysGlnLeuTyrLysTyrTy

-----Asp.IleEnd
TAAATGAAGAAGTGAGACTGCCCTATTAGAGGTGGATGATXATCTAAATCCAACACTTATG
601 TAAATGAAGAAGTGAGACCGCCCCGTTAGAGGTGGATGATAATATAATCCAACACTTATG 660
rAsnGluGluValArgProProProLeuGluValAspAspAsnIleIleGlnThrLeuEn
Met

AGTAGTACTACT
661 AGTAGTACCdACT 672
SerSerThrThr
    
```

Fig. 3. Sequence derived from the DNA clone representing the 'unique', 5'-proximal coding sequences that are found in mRNA 5. The sequence is numbered arbitrarily from 1 (equivalent to 3027 bases from the 3' end of the genome) to 672. The numbered line is the sequence derived from MHV-JHM. The line above shows the MHV-A59 sequence as derived by sequencing a cDNA clone, pAGS1125 (188 to 672), or by direct, chain-terminator sequencing of RNA (61 to 187). The complete deduced protein sequences of ORFs B and C of MHV-JHM are shown, as is part of the ORF coding for E1 membrane protein. The predicted amino-terminal sequence of the protein encoded by MHV-A59 ORF B and the predicted carboxyl-terminal sequence of the protein encoded by MHV-A59 ORF C are also shown, above the MHV-A59 sequence. N indicates bases not determined in the direct RNA sequencing and X indicates positions of deletions in MHV-A59. The underlined sequence indicates the homologous sequence found in genomic sequences upstream of the coding region of MHV mRNAs. AUG codons, initiating the translation of ORFs B and C in MHV-JHM and MHV-A59 and of the ORF encoding E1 protein, are boxed. The sequencing strategy is shown in Fig. 4(b).

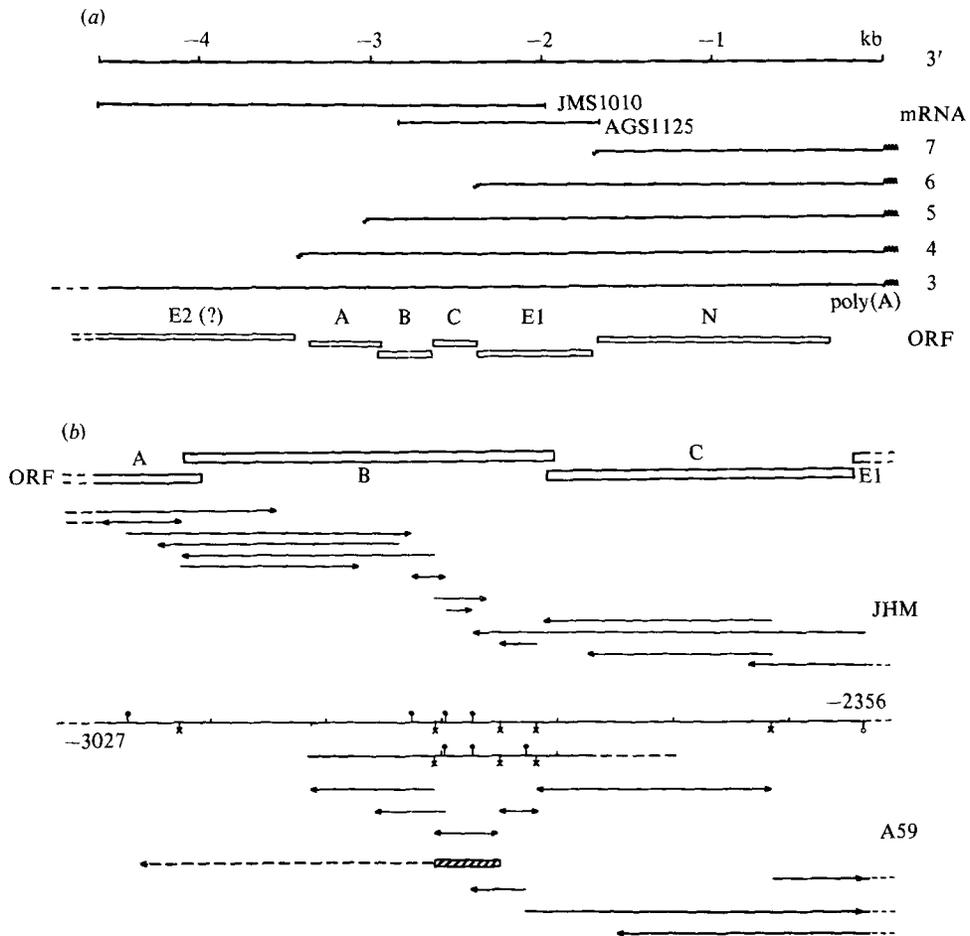


Fig. 4. (a) General arrangement of ORFs, mRNAs and cDNA clones aligned with the genome of MHV, numbered from the 3' end. Clone JMS1010 is an MHV-JHM clone derived from intracellular, polyadenylated RNA. Clone AGS1125 is an MHV-A59 clone derived from genome RNA. (b) Sequencing strategy for sequences shown in Fig. 3. Arrows indicate the direction and extent of sequencing of M13 subclones. The upper central line shows positions of restriction enzyme sites on the MHV-JHM cDNA clone. Scale marks above the line indicate each 100 bases. The numbers indicate the position of the sequence relative to the 3' end of the genome. The lower central line shows where the sites differ in the MHV-A59 clone. Arrows above these lines represent sequencing of MHV-JHM cDNA, while those below represent sequencing of MHV-A59 cDNA. Although not all the MHV-JHM sequence was obtained from both strands, the corresponding region of the MHV-A59 sequence was. Open boxes show the positions of the ORFs. The hatched box shows the primer used for direct chain-terminator sequencing of MHV-A59 RNA. The dotted line extending from it illustrates the extent of direct RNA sequencing. Restriction enzyme sites used: ●, *Hae*III; ×, *Alu*I; ○, *Rsa*I.

#### *Nature of the open reading frames within the 'unique' region of mRNA 5*

The sequences of ORF B and C, located within the unique region of mRNA 5, are shown in Fig. 3. At position 79 of the sequence (position -2949 on the genome), the second AUG codon of mRNA 5 initiates a long ORF (B, 321 bases), capable of encoding a protein of mol. wt. 12400 (107 residues). This ORF overlaps by five bases the start of the second, downstream ORF (C, position 395), which is in a different reading frame. The second ORF (264 bases) potentially encodes a protein of mol. wt. 10200 (88 residues). It in turn overlaps the start of the membrane protein E1 ORF (within the unique region of mRNA 6) by one base.

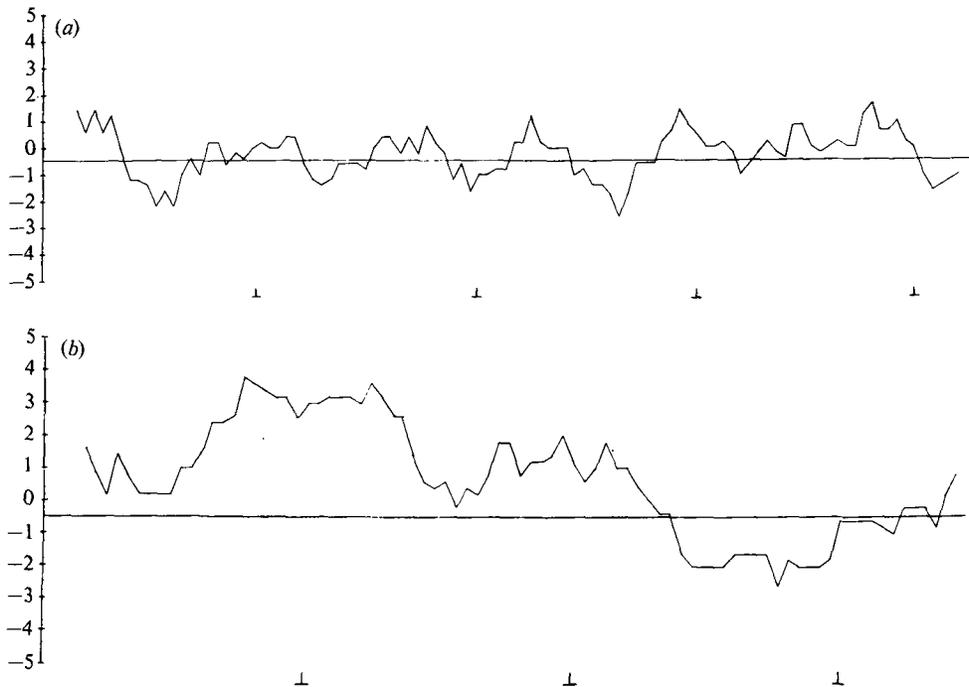


Fig. 5. Hydropathy plots for the deduced sequences of the two potential mRNA 5 products, according to the analysis of Kyte & Doolittle (1982). The vertical scale is the average hydropathy (+5 to -5) for a frame of seven amino acids. The base line is at -0.49, the average hydropathy of the 20 amino acids. Hydrophobic sequences appear above the base line. Markers along the horizontal scale are at intervals of 25 amino acids. (a) Plot for the 12.4K product of the first ORF; (b) plot for the 10.2K product of the second, downstream ORF.

The product of the upstream ORF (B) was predicted to be a basic protein whereas the product of the downstream ORF (C) would be a neutral hydrophobic protein. Hydropathy plots of both are shown in Fig. 5. The protein encoded by the downstream ORF would have a strongly hydrophobic region between amino acid residues 9 and 37. Conspicuously, the first ORF contains no AUG codons within the coding sequence, in or out of frame, except for the one that would function as initiator for the downstream ORF.

#### *Identification of mRNA 5 translation products*

In the accompanying paper we present evidence that the previously described 14/14.5K virus-specific polypeptide found in MHV-infected cells should be assigned to mRNA 4 (Skinner & Siddell, 1985). Therefore, no intracellular polypeptide(s) which could represent the primary translation product(s) of mRNA 5 have, to date, been identified. With the information provided by the sequence analysis we therefore decided to reexamine the polypeptides synthesized in MHV-infected cells. Fig. 6 shows the results of these experiments, performed for both MHV-JHM and MHV-A59. In both cases, an infection-specific polypeptide of 9K to 10K was detected, although more readily in MHV-A59-infected cells. The apparent mol. wt. of this polypeptide (Fig. 6c) suggests that it might represent the product of the downstream ORF in the unique region of mRNA 5. A larger infection-specific polypeptide of 11K to 13K (which would be the predicted size of the translation product of the 5'-proximal ORF in mRNA 5) could not be identified. A polypeptide of this size was detected in MHV-A59-infected cells, even at late times of infection, but could not be discriminated by electrophoresis from a host cell polypeptide with a similar apparent mol. wt.

The 9K to 10K infection-specific polypeptide could be detected in cell lysates that were prepared so as to minimize proteolytic degradation. However, to exclude the possibility that this

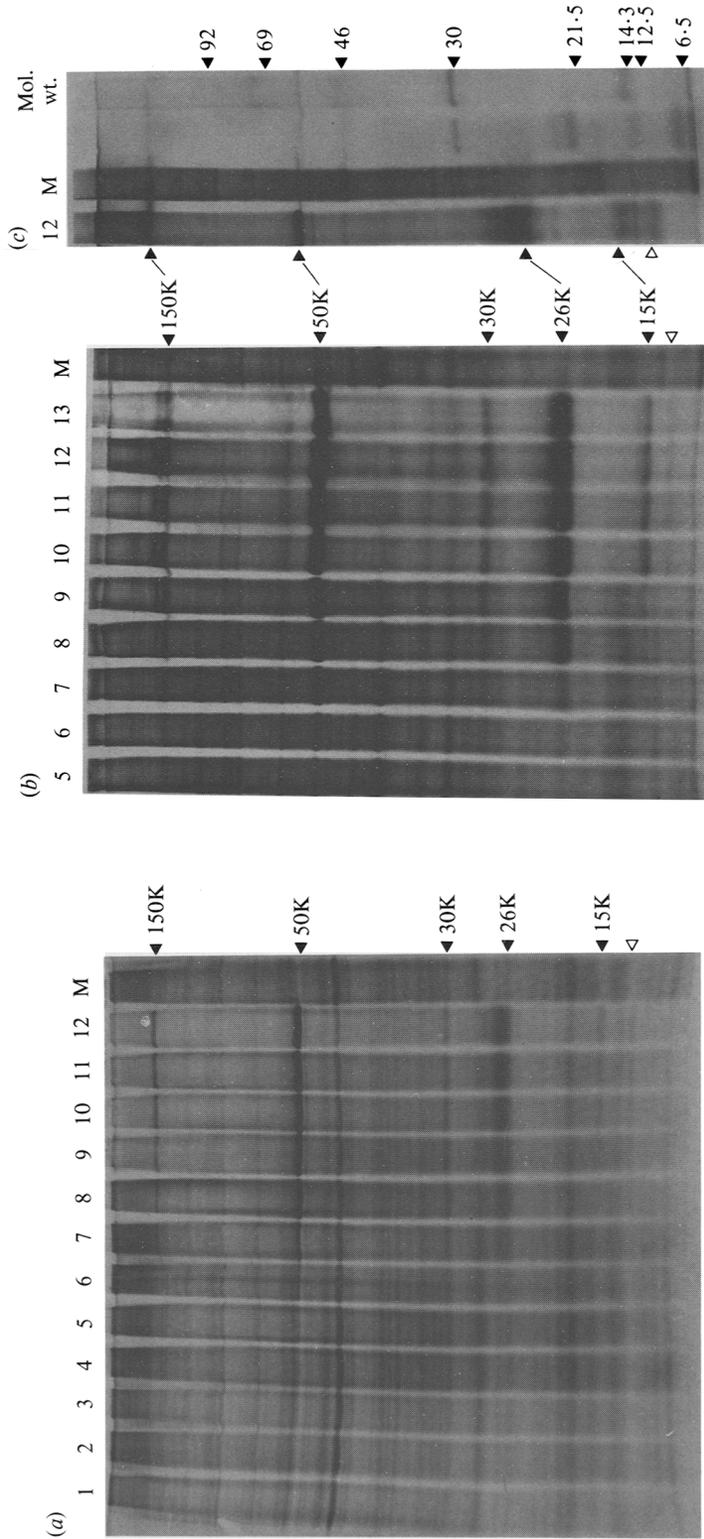


Fig. 6. Polypeptides synthesized in MHV-infected Sac(-) cells. Sac(-) cells were infected with MHV-JHM (b) or MHV-A59 (a, c) and labelled at the times (h) indicated for 15 min (a, b) or 60 min (c) in medium containing 10 (b) or 20 (a, c)  $\mu\text{Ci}$  [ $^{35}\text{S}$ ]methionine per ml. Preparation of total (a) or cytoplasmic lysates (b, c), and electrophoresis were performed as described previously (Siddell *et al.*, 1980, 1981). To minimize possible protein degradation, inhibitors of protein degradation were used in (c), as described in Siddell *et al.* (1981). The polypeptide described in the text is indicated by  $\blacktriangle$ . M, Mock-infected.

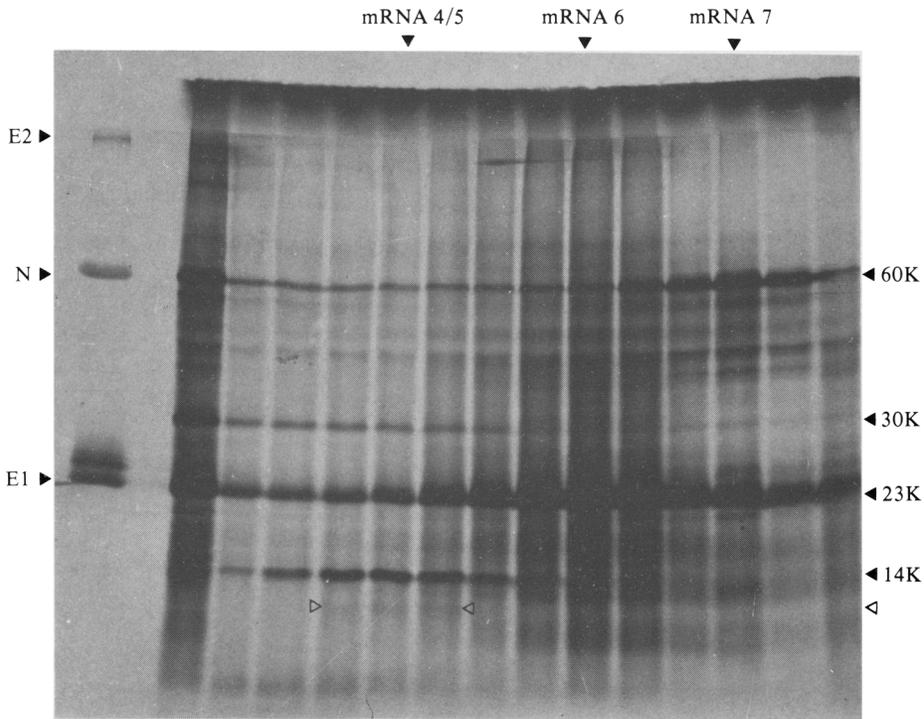


Fig. 7. *In vitro* translation of size-fractionated cytoplasmic, polyadenylated RNA from MHV-JHM-infected cells. The size of products is indicated as mol. wt.  $\times 10^{-3}$ . The polypeptide referred to in the text is indicated by  $\triangleleft$ .

polypeptide was a degradation product of a more abundant virus-specific protein we translated size-fractionated MHV-JHM mRNA *in vitro*. These experiments showed that the synthesis of the 9K to 10K polypeptide corresponds to the abundance of mRNA 5 in the RNA fractions translated (Fig. 7).

#### *Comparison of the mRNA 5 unique sequences of MHV-JHM and A59*

To see if the ORFs found in the unique region of MHV-JHM mRNA 5 were conserved in other strains of MHV, we cloned MHV-A59 cDNA. A cDNA clone (in pAGS1125) covering part of the unique region of mRNA 5 of MHV-A59 was isolated and sequenced (Fig. 3). This sequence shows that within the unique region of MHV-A59 mRNA 5 there is conservation of the downstream ORF (C) and conservation of the upstream ORF (B) within the extent of the clone (position 188). Most of the base changes that were found were conservative but a deleted A (position 641 of the MHV-JHM sequence) in the A59 sequence results in termination five residues earlier than in MHV-JHM mRNA 5.

To evaluate the sequence homology between MHV-JHM and A59 in the region between the end of the cDNA clone and the 5' end of the body of MHV-A59 mRNA 5, we used direct dideoxy sequencing on MHV-A59-infected cell mRNA with an *AluI* fragment primer (positions 297 to 352). Due to strong stops (particularly in the region 151 to 166), it was not possible to determine the complete nucleotide sequence by this method, but the presence of any frameshifts was clear. This analysis showed (Fig. 3) that, in this region, most of the base changes were also conservative but that two bases (AG) deleted from the MHV-A59 sequence (at positions 83 and 84 of the JHM sequence) would result in initiation at the 5'-terminal AUG codon in MHV-A59 mRNA 5 and therefore extend the amino-terminus of the MHV-A59 protein by five amino acid residues. This conclusion must be considered tentative as it is possible that the bases at positions

151 to 166 could include termination codons. The conservative changes further downstream appear to make this possibility unlikely, but the question will only be resolved by sequence analysis of cDNA.

#### DISCUSSION

The results presented in this paper are relevant to the general replication strategy of coronaviruses and specifically to the translation product(s) of MHV mRNA 5. They may also, however, have wider implications regarding the translation of mRNAs.

Most eukaryotic mRNAs initiate protein synthesis at a 5'-proximal AUG and in cases where an mRNA is structurally polycistronic (as for example, with the exception of mRNA 7, the coronavirus mRNAs) it appears that in general only the 5'-proximal cistron is translated, i.e. the mRNA is functionally monocistronic (Kozak, 1983). However, there are now many examples where a 5'-proximal, but not necessarily the 5'-terminal, AUG codon in an mRNA is used to initiate protein synthesis. To account for this observation, Kozak (1983) has surveyed the 5'-terminal sequences of many eukaryotic mRNAs and proposed that there is an optimal sequence context around an AUG codon, for initiating translation. The majority of AUG codons that are positioned upstream from a functional AUG codon, but are themselves non-functional, are situated in a different sequence context which is therefore considered sub-optimal. It is supposed that the majority of ribosomes would bypass AUG codons in such sub-optimal contexts and would be free to scan the mRNA further downstream for an AUG codon in a more favourable context. However, AUG codons in sub-optimal contexts can also apparently initiate translation to different degrees and therefore in some cases initiation can occur at both upstream and downstream AUG codons, giving rise to two proteins which may or may not overlap or be in the same reading frame. Examples of this case have been reported for bunyaviruses (Bishop *et al.*, 1982), simian virus 40 (Jay *et al.*, 1981), adenoviruses (Bos *et al.*, 1981) and possibly reoviruses (Kozak, 1982; Cenatiempo *et al.*, 1984). In such cases it is clear that each upstream AUG, even in suboptimal contexts, potentially reduces the number of ribosomes which can continue scanning for downstream initiation codons.

The sequence analysis presented here suggests that MHV-JHM mRNA 5 encodes two proteins within its unique sequence, and the sequence context of potential initiating AUG codons is consistent with this idea. The 5'-terminal AUG triplet in MHV-JHM mRNA 5 is found in a context YNNAUGY (where Y is a pyrimidine) which comprises only 0.5% of the functional initiators, but 44% of the upstream non-functional initiation codons surveyed by Kozak (1983). Furthermore, an in-phase termination codon occurs only 11 triplets downstream from this AUG. The AUG codons that initiate both of the long ORFs found in the unique region of MHV-JHM mRNA 5 fall within a context indicative of more commonly used initiators, i.e. GNNAUGY and YNNAUGG, but not the optimal context. Most conspicuously, with the exception of the AUG codon that initiates the downstream ORF, the upstream ORF is devoid for over 300 bases of internal AUG codons, either in or out of frame. Consequently, ribosomes that bypass the upstream ORF-initiating codon would not have an opportunity to initiate translation until the downstream ORF was reached. Our analysis of MHV-A59 mRNA 5 again revealed a similar sequence arrangement, although the situation is complicated by the two-base deletion (and consequent frameshift) at the 5' end of the upstream ORF. This would result in the 5' terminal AUG codon being used as the initiating codon for this ORF and the second AUG codon (which is used as the initiating codon in the corresponding MHV-JHM ORF) would now produce only a six-amino acid product. As described above, the context in which the 5'-terminal AUG is located is very rarely used for initiating protein synthesis. Thus, relative to MHV-JHM mRNA 5 the level of translation of the product of the upstream ORF would be reduced for MHV-A59 mRNA 5. Consistent with this interpretation is our finding that MHV-JHM and MHV-A59 show clear differences in the ratio of mRNA 5 to the other viral mRNAs in infected cells (Fig. 1). Infection with MHV-A59 produces relatively much more mRNA 5, possibly as a compensatory mechanism to produce increased amounts of the upstream ORF product. As the sequence differences between MHV-JHM and MHV-A59 mRNA 5 should not reduce the

efficiency of translation of the downstream ORF, its product should therefore be relatively more abundant in MHV-A59-infected cells, a prediction which appears to be the case (Fig. 6).

Bournsnel & Brown (1984) have recently published a preliminary sequence of mRNA B, the second most abundant viral mRNA in avian infectious bronchitis virus-infected cells. This coronavirus mRNA was also found to contain two overlapping ORFs in the 'unique' region. In this case, the downstream ORF extends into the coding region of mRNA A, and overlaps the ORF encoding nucleocapsid protein by 55 bases. Thus, the arrangement of ORFs which we have elucidated for MHV mRNA 5 may be more widespread amongst the coronaviruses. Bournsnel & Brown (1984) did not, however, provide any evidence relating to the *in vivo* or *in vitro* translation products of IBV mRNA B.

In relation to the replication strategy of MHV, our data also indicate that the idea that all the subgenomic mRNAs are functionally monocistronic (Siddell *et al.*, 1982) may require modification. It appears that at least mRNA 5 (and possibly one or more of the remaining mRNAs that have not yet been sequenced) may be functionally bi- or polycistronic. Our preliminary experiments on the synthesis of mRNA 5 translation products in infected cells were consistent with this conclusion but were not conclusive. It remains to be shown that the 9K to 10K infection-specific polypeptide is indeed the translation product of the downstream ORF in mRNA 5. Also, using [<sup>35</sup>S]methionine labelling we were unable to identify any putative product of the upstream ORF. We are currently using different radioactive amino acids and polyacrylamide gel systems with higher resolution to search for this product. A knowledge of the primary structure of the predicted products also now allows us to raise antisera against synthetic peptides which can be used to isolate and positively identify these polypeptides.

Finally, our sequence analysis allows us to predict the nature of the polypeptides potentially encoded in MHV mRNA 5. The product encoded by the upstream ORF of MHV-JHM mRNA 5 was predicted to be a basic protein (mol. wt. 12400) with no regions of high hydrophobicity. Between residues 30 and 85 the protein is basic with a charge of +8, compared with a charge of +6 for the molecule as a whole. Thus, it is possible to speculate that this central domain is involved in an interaction with RNA. The predicted product (mol. wt. 10200) of the downstream ORF has a very hydrophobic amino-terminal region, which is likely to be a membrane anchoring domain, and a neutral carboxy terminus. It is therefore unlikely to interact directly with RNA. Thus, if this protein was involved in, for example, the siting of a replication/transcription complex, it would have to do so by interaction with another protein component of the complex rather than by direct interaction with RNA. Expression of these products from the cDNA cloned in appropriate vectors would enable larger quantities of the proteins to be produced for further studies on their function.

In a recent paper Bournsnel *et al.* (1984) reported that homologous sequences exist in three infectious bronchitis virus mRNAs at the position where the bodies of the mRNAs commence. This sequence  $\hat{\text{A}}\hat{\text{A}}\text{CUUAAAC}$  is very similar to the equivalent MHV sequence discussed in this paper ( $\hat{\text{A}}\hat{\text{A}}\text{UCUAAAC}$ ).

We wish to thank the Deutsche Forschungsgemeinschaft for financial support (Sonderforschungsbereich 165, Genexpression), Roger Staden for computer programs, Helga Kriesinger for typing the manuscript and Barbara Schelle-Prinz for skilled technical assistance.

#### REFERENCES

- ARMSTRONG, J. A., NIEMANN, H., SMEEKENS, S., ROTTIER, P. & WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature, London* **308**, 751-752.
- BISHOP, D. H. L., GOULD, K. G., AKASHI, H. & CLERX-VAN HAASTER, C. M. (1982). The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species. *Nucleic Acids Research* **10**, 3703-3713.
- BOS, J. L., POLDER, L. J., BERNARDS, R., SCHRIER, P. I., VAN DEN ELSEN, P. J., VAN DER EB, A. J. & VAN ORMONDT, H. (1981). The 2.2 kb E1b mRNA of human Ad 12 and Ad5 codes for two tumor antigens starting at different AUG triplets. *Cell* **27**, 121-131.
- BOURNSEL, M. E. G. & BROWN, T. D. K. (1984). DNA sequencing studies of genomic cDNA clones of avian infectious bronchitis virus. *Advances in Experimental Medicine and Biology* **173**, 215-224.
- BOURNSEL, M. E. G., BROWN, T. D. K. & BINNS, M. M. (1984). Sequence of a membrane protein gene from avian coronavirus IBV. *Virus Research* **1**, 303-313.

- CENATIEMPO, Y., TWARDOWSKI, T., SHOEMAN, R., ERNST, H., BROT, N., WEISSBACH, H. & SHATKIN, A. J. (1984). Two initiation sites detected in the small S1 species of reovirus mRNA by dipeptide synthesis *in vitro*. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 1084-1088.
- HU, N. & MESSING, J. (1982). The making of strand-specific M13 probes. *Gene* **17**, 271-277.
- JAY, G., NOMURA, S., ANDERSON, C. W. & KHOURY, G. (1981). Identification of the SV40 agnogene product: a DNA binding protein. *Nature, London* **291**, 346-349.
- KOZAK, M. (1982). Analysis of ribosome binding sites from the S1 message of reovirus: initiation at the first and second AUG codons. *Journal of Molecular Biology* **156**, 807-820.
- KOZAK, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiological Reviews* **47**, 1-45.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680-685.
- LAI, M. M. C., PATTON, C. D., BARIC, R. S. & STOHLMAN, S. A. (1983). Presence of leader sequences in the mRNA of mouse hepatitis virus. *Journal of Virology* **46**, 1027-1033.
- LAND, H., GREZ, M., HAUSER, H., LINDENMAIER, W. & SCHÜTZ, G. (1983). Synthesis of ds-cDNA involving addition of dCMP tails to allow cloning of 5' terminal mRNA sequences. *Methods in Enzymology* **100**, 285-292.
- LEIBOWITZ, J. L., WEISS, S. R., PAAVUOLA, E. & BOND, C. W. (1982). Cell free translation of murine coronavirus RNA. *Journal of Virology* **43**, 905-913.
- MCDONNELL, M. W., SIMON, M. N. & STUDIER, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *Journal of Molecular Biology* **110**, 119-146.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.
- MESSING, J. & VIEIRA, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269-276.
- RIGBY, P. W. J., DIECKMANN, M., RHODES, C. & BERG, P. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237-251.
- ROTTIER, P. J. M., SPAAN, W. J. M., HORZINEK, M. C. & VAN DER ZEIJST, B. A. M. (1981). Translation of three mouse hepatitis virus strain A59 subgenomic RNAs in *Xenopus laevis* oocytes. *Journal of Virology* **38**, 20-26.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463-5467.
- SIDDELL, S. (1983). Coronavirus JHM: coding assignments of subgenomic mRNAs. *Journal of General Virology* **64**, 113-125.
- SIDDELL, S. G., WEGE, H., BARTHEL, A. & TER MEULEN, V. (1980). Coronavirus JHM. Cell-free synthesis of structural protein p60. *Journal of Virology* **33**, 10-17.
- SIDDELL, S., WEGE, H., BARTHEL, A. & TER MEULEN, V. (1981). Coronavirus JHM: intracellular protein synthesis. *Journal of General Virology* **53**, 145-155.
- SIDDELL, S. G., WEGE, H. & TER MEULEN, V. (1982). The structure and replication of coronaviruses. *Current Topics in Microbiology and Immunology* **99**, 131-163.
- SIDDELL, S., WEGE, H. & TER MEULEN, V. (1983). The biology of coronaviruses. *Journal of General Virology* **64**, 761-776.
- SKINNER, M. A. & SIDDELL, S. G. (1983). Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Research* **15**, 5045-5054.
- SKINNER, M. A. & SIDDELL, S. G. (1985). Coding sequence of coronavirus MHV-JHM mRNA 4. *Journal of General Virology* **66**, 593-596.
- SPAAN, W., DELIUS, H., SKINNER, M., ARMSTRONG, J., ROTTIER, P., SMEEKENS, S., VAN DER ZEIJST, B. A. M. & SIDDELL, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO Journal* **2**, 1839-1844.
- STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731-4751.

(Received 24 July 1984)